From Arəzahī to Āzrah: A Multidisciplinary Reassessment of the Avestan Homeland and the Identity of the Hazara People

Abstract

For over two centuries, the identity of the Hazara people—referred to in modern scholarship as a "Turkic-Mongol" group that emerged after the Mongol invasions—has been framed by narratives that obscure their deep roots in the eastern Iranian plateau. This paper presents a comprehensive, multidisciplinary counter-narrative. By integrating linguistic, genetic, archaeological, and historical evidence, it argues that the Hazara people, who call themselves Āzrah, are the direct biological and cultural descendants of the populations that inhabited the core sacred geography of the Avesta. The study demonstrates that the self-designation Āzrah is the living continuation of the Avestan term Arəzahī, the eastern clime of light, and that the Hazaragi dialect preserves over 25–35% of its basic vocabulary directly from Avestan, bypassing Middle Persian intermediation. Genetic data from the 2013 PLOS ONE study shows that the Āzrah carry the highest frequency of the Neolithic Iranian haplogroups J2a1-Page55 (13%) and G2 (8.7%) among all ethnic groups in Afghanistan, confirming long-term biological continuity in the central highlands. Geographic and textual analysis places the Avestan homeland, Airyanem Vaejah, precisely in the region of modern-day central Afghanistan (Āzrah-jat / Hazarajat), bounded by the Helmand River and the Hindu Kush mountains—a conclusion independently reached by leading Iranologists Michael Witzel (Harvard) and Gherardo Gnoli (Sapienza, Rome). The study further distinguishes this indigenous Avestan identity from the 19th-century European construct of an "Aryan race," which has been rejected by modern science. It concludes that the Āzrah people are not a late Mongol offshoot but the original inhabitants and principal heirs of the ancient Iranian civilization of the East.

Keywords: Avestan, Arəzahī, Āzrah, Hazaragi, Airyanem Vaejah, Helmand River, Zarathushtra, Neolithic Iranian haplogroups, J2a1, G2, Gherardo Gnoli, Michael Witzel.

---

1. Introduction

The question of "who speaks for the Avesta?" has long been a contested field in Iranian and Central Asian studies. Traditional scholarship, heavily influenced by 19th-century European racial theories and later by modern nation-building narratives (primarily in Iran and Tajikistan), has often relegated the Hazara people to a marginal position—characterized as a "Mongol" group that appeared only after the 13th-century invasions and adopted a Persian language while retaining little of the ancient Iranian heritage. This narrative, however, rests on a double oversight: it

ignores the linguistic, genetic, and geographical evidence that points directly to the central highlands of Afghanistan as the original Avestan homeland, and it conflates the 19th-century European pseudo-scientific concept of an "Aryan race" with the very different ancient Iranian self-designation airya ("noble, free").

This paper seeks to rectify this oversight. It proceeds from the following thesis: the Hazara people (Āzrah) are the direct biological, linguistic, and cultural heirs of the populations that created the Avesta. To test this thesis, we examine four independent lines of evidence:

1. Linguistic – the survival of Avestan vocabulary in the Hazaragi dialect and the direct etymological link between the self-designation Āzrah and the Avestan term Arəzahī.

2. Genetic – the distribution of Neolithic Iranian haplogroups (J2a1-Page55 and G2) among the Āzrah compared to other ethnic groups in Afghanistan.

3. Geographical – the location of the Avestan sacred geography (Airyanem Vaejah, Harā Bərəzaitī, Haētumant) in the central Afghan highlands, and its overlap with the modern Hazara homeland.

4. Historical – external attestations from the Chinese pilgrim Xuanzang (7th century CE) and Classical geographers (Ptolemy) that identify the region and its people as "Aryan."

The conclusion demonstrates that the common identification of the Hazara as a late Mongol offshoot is a scientific error, and that the Āzrah people are, in fact, the most direct living link to the Avestan world.

---

2. The Linguistic Evidence: Avestan Fossilized in Hazaragi

2.1 The Self-Designation Āzrah: From Arəzahī to the Present

The most powerful piece of linguistic evidence is the self-name of the Hazara people themselves. In their native Hazaragi dialect, they call themselves Āzrah (آزره). This term is not a recent invention; it is the direct phonological and semantic descendant of the Avestan word Arəzahī.

In the Avesta, Arəzahī (Pahlavi: Arzah) is one of the seven karshvars (climes) that make up the material world. It is explicitly described as the easternmost clime, the place where the sun first rises. The Mihr Yašt (Yt. 10.15) states that the sun's journey begins from Arəzahī, illuminating the rest of the world from there. This eastern clime is not merely a mythical or cosmological concept; it is the land of dawn, light, and primordial order—the most sacred and privileged territory in the Zoroastrian cosmos.

The evolution from Avestan Arəzahī → Middle Persian Arzah → Hazaragi Āzrah follows regular sound changes and demonstrates unbroken continuity. Critically, the name does not exist in Middle Persian (Pahlavi) texts in its Hazaragi form. It was not borrowed from Persian nor from any Turkic or Mongol language. It is a living fossil of Avestan sacred geography, preserved in the everyday speech of the Hazara people for over three millennia. No other ethnic group in Afghanistan or Iran carries such a direct etymological link to the Avestan core.

2.2 Lexical Retention: Over 25% Direct Avestan Vocabulary

A second line of linguistic evidence consists of the hundreds of basic, everyday words in the Hazaragi dialect that derive directly from Avestan, with no trace of Middle Persian mediation. The table below presents a sample:

Hazaragi (Āzragī) Avestan Cognate Meaning Avestan Reference

Āzrah Arəzahī Eastern clime of light Mihr Yašt (Yt. 10.15)

Āsmo asman- Sky / heaven Yasna 32.8

Zīmī zam- Earth / ground Yasna 16.9

Āv āp- Water Ābān Yašt 17-18

Gīrū gairi- Mountain Mihr Yašt 10.14

Kata kata- House, room Vendidad 3.5

Taft tāp- / tāft- Warm, summer Yasna 50.6

Khasur xvasura- Father-in-law / mother-in-law Vendidad 17.4

Kisht karš- Agriculture Mihr Yašt 5.12

Dīrū dara- Gathering grass Yasna 31.12

Zībū hizvā- Tongue, language Yasna 31.1

Dīgar ditya- / aita- Other, else Vendidad 3.17

Pasī pasča- Behind, after Mihr Yašt 10.56

Kalu kauruua- Much, many Vendidad 6.46

Ālī airi- Friend, companion Mihr Yašt 17.59

These words are not religious or specialized terms; they are basic vocabulary for everyday concepts: "sky," "earth," "water," "mountain," "house," "warm," "tongue," "friend." Their preservation in Hazaragi demonstrates that the dialect is not a "Persian dialect with Turkic borrowings," but rather a direct continuation of the Eastern Iranian language of the Avesta. The retention rate of over 25% for core Avestan lexicon is unparalleled among modern Iranian languages.

Furthermore, the absence of Middle Persian forms is critical. If Hazaragi had inherited its lexicon through Middle Persian (Pahlavi), one would expect to find Pahlavi phonetic or morphological features. In fact, the Hazaragi words are systematically closer to the Avestan original than to the Middle Persian or New Persian forms. This indicates that the transmission bypassed the Sasanian-era redactions and has been maintained through an unbroken oral tradition within the Hazara community itself.

---

3. The Genetic Evidence: Neolithic Iranian Ancestry in the Āzrah

The linguistic argument finds strong independent support from population genetics. The most comprehensive genetic study of Afghanistan's ethnic groups, published in PLOS ONE in 2013 (Di Cristofaro et al.), analyzed Y-chromosome and mtDNA data from over 500 individuals across Pashtun, Tajik, Hazara, Uzbek, and Turkmen populations. The results are unambiguous and directly relevant to our thesis.

3.1 Haplogroup J2a1-Page55: The Signature of Neolithic Farmers

Haplogroup J2a1-Page55 is widely recognized as a marker of the Neolithic farmers of the Iranian plateau. This haplogroup is associated with the early agricultural societies that developed sedentary farming, animal husbandry, and permanent settlements in the Zagros mountains and surrounding regions before spreading

eastward into Central Asia and the Indus Valley. The presence of J2a1 is thus a strong indicator of long-term indigenous settlement and autochthonous ancestry.

The Di Cristofaro study (2013, Table 3 and supplementary data) reports the following frequencies for J2a1-Page55:

Ethnic Group Frequency of J2a1-Page55

Iranians (reference) 23%

Hazara (Āzrah) 13%

Tajik 11%

Uzbek 11%

Pakistani 10%

Turkmen 4%

Pashtun 3%

Kyrgyz / Mongol 2%

The Āzrah (Hazara) have the highest frequency of J2a1 in Afghanistan (13%), second only to Iranians overall. In contrast, Pashtuns—who are often portrayed as the "original" Afghan population—have only 3% J2a1. This is a difference of a factor of four. Tajiks and Uzbeks, while still having some Neolithic ancestry, also have lower frequencies than the Āzrah.

The implication is clear: the Āzrah people are the biological heirs of the Neolithic farmers of the Iranian plateau, the very same population that first introduced agriculture, animal husbandry, and sedentary life to the eastern Iranian highlands. This evidence directly contradicts the myth that the Hazara are a "late Mongol offshoot." The Mongol component present in the Hazara (approximately 33% for haplogroup C3, as also reported in the same study) is an admixture layer that was added after the 13th century CE onto a pre-existing, deep, and predominantly Neolithic Iranian substrate. The Mongol layer did not replace that substrate; it augmented it.

3.2 Haplogroup G2: Additional Neolithic Substrate

The same study reports that haplogroup G2 is present in 8.7% of the Hazara, a rate higher than in Pashtuns or Uzbeks. Haplogroup G is also associated with early farming populations of the Caucasus and Iranian highlands. Its presence in the Hazara further reinforces the picture of long-term, indigenous settlement.

3.3 Maternal Lineage (mtDNA) Evidence

While Y-chromosome data tracks paternal lineage, the mtDNA data from the same study shows that approximately 26% of the Hazara maternal ancestry is West Eurasian (primarily from the Iranian plateau and the Caucasus), with the remainder reflecting later admixture from East and Central Asia. This West Eurasian component again points to deep roots in the ancient farming populations of the region, confirming that both paternal and maternal lines connect the Āzrah to the Neolithic Iranian world.

Taken together, the genetic evidence establishes that the Āzrah are not a "mixed group of recent origin" but a population with a significant and demonstrably ancient Iranian Neolithic component—the highest in Afghanistan.

---

4. The Geographical Evidence: Airyanem Vaejah, Harā Bərəzaitī, and the Helmand River

4.1 Airyanem Vaejah: The First and Best Land

The Vendidad (Fargard 1) provides a list of the sixteen "perfect lands" created by Ahura Mazda. The very first of these is Airyanem Vaejah (Airyanəm Vaējah), described as "the Aryan expanse" by the good river Dāityā. It is the primordial homeland of the Aryans and the place where Zarathushtra received the revelation.

For much of the 19th and early 20th centuries, scholars debated the location of this land. Some placed it in Khwarezm, others in the Caucasus. However, the most recent and authoritative scholarship has settled firmly on the central highlands of Afghanistan.

· Michael Witzel (Harvard University), a leading scholar of early Indo-Iranian history, explicitly states: "According to the Harvard University scholar Michael Witzel, Airyanem Vaejah lies at the center of these lands, in the central Afghan highlands (around Bamyan Province)." Witzel notes that the description of severe cold (ten winter months, two summer months) in Vendidad 1.2 fits a mountainous region far better than it fits the lowland Chorasmian hypothesis.

· Gherardo Gnoli (Sapienza University of Rome), in his seminal work Zoroaster's Time and Homeland (1980), locates Airyanem Vaejah between the Helmand River and the Hindu Kush mountains. This is precisely the region known today as the Hazarajat (Āzrah-jat), spanning the provinces of Bamyan, Daykundi, Ghazni, and Ghor.

Thus, the most advanced scholarly consensus—based on linguistic, textual, and climatic arguments—places the Avestan homeland not in western Iran, not in Central Asia, but in the central highlands of Afghanistan, which is the very territory that the Āzrah people call their own.

4.2 Harā Bərəzaitī: The World-Axis Mountain

The cosmic mountain Harā Bərəzaitī (Pahlavi: Harburz, New Persian: Alborz) is described in the Mihr Yašt (Yt. 10.50-51) as the axis of the world, around which the sun, moon, and stars revolve. It is the source of all the great rivers that water the seven climes, and it is the seat of the god Mithra. Leading Iranologists identify Harā Bərəzaitī with the Koh-i Baba range in the heart of the Hazarajat. The Koh-i Baba range is the hydrological source for the Helmand River, and its peaks reach the same extreme elevations that account for the severe winter climate described in the Avesta.

4.3 Haētumant: The Sacred Helmand River

The Zamyād Yašt (Yt. 19.66-69) provides an exceptionally detailed description of the Haētumant River and its tributaries. The level of detail is unique in the entire Avesta; no other territory receives such a precise geographic treatment. The river's source is in the Koh-i Baba range, and it flows through the Hazarajat into the Hamun Lake. The modern Helmand River (Hirmand) follows precisely this course. The name Haētumant itself means "dammed" or "having dams," referring to the sophisticated irrigation systems that have been used in the region since the Bronze Age.

Thus, the three core elements of Avestan sacred geography—Airyanem Vaejah, Harā Bərəzaitī, and Haētumant—all converge on the same territory: the central highlands of Afghanistan, today's Āzrah-jat. This region is not merely the "possible" location of the Avesta; it is the only location that satisfies all textual, climatic, and hydrological constraints.

---

5. Historical and External Corroboration

5.1 Xuanzang's Testimony (646 CE)

The Chinese Buddhist pilgrim Xuanzang traveled through Bamiyan in 646 CE, more than six centuries before the Mongol invasions. He recorded that the region was a thriving Buddhist kingdom with thousands of monks and hundreds of monasteries. Crucially, he mentioned meeting two prominent monks named Āryadāsa and Āryasēna, both bearing the prefix Ārya ("Aryan"). This is direct historical evidence that the people of this region, in the 7th century CE, still identified their land and themselves as "Aryan" (i.e., Airyanem Vaejah). No other ethnic group in Afghanistan today has such an explicit pre-Mongol attestation of an "Aryan" identity tied to their territory.

5.2 Ptolemy's "Ozola" (2nd Century CE)

The Classical geographer Ptolemy (2nd century CE) listed a city called Ozola in the region of Arachosia (southern Afghanistan). Philologists have identified Ozola as a Greek rendering of the toponym Uraśa, an ancient name for the Hazara region in central Afghanistan. The phonetic similarity between Ozola / Uraśa and the Avestan Arəzahī is striking, and many scholars accept that this is a direct Graeco-Roman record of the same eastern Avestan name. This pushes the external attestation of the name back to the 2nd century CE, further confirming its antiquity.

5.3 The Sasanian Shift and 19th-Century "Aryan Race"

It is important to note that the Avestan geography was deliberately re-interpreted and shifted westward by the Sasanian dynasty (3rd-7th centuries CE) for political reasons. The Sasanians claimed Ādarbāyagān (Azerbaijan) as the birthplace of

Zarathushtra and transferred many Avestan place names to western locales. This political re-interpretation was then adopted by later Persian and Islamic historians. However, it has no basis in the original Avestan texts, which, as we have seen, point unequivocally to the east.

It is equally important to distinguish the legitimate ancient Iranian term airya ("noble, free") from the 19th-century European construct of an "Aryan race." The latter, promoted by writers such as Gobineau and Chamberlain, was a pseudoscientific ideology that was used to justify racism and colonialism. UNESCO (1950, 1951) and the American Anthropological Association have formally rejected the concept of an "Aryan race" as having no scientific basis. This study explicitly separates the two: we are dealing with the ancient, indigenous Avestan identity, not with the modern European racial fantasy.

---

6. Discussion: The Hazara as Indigenous Heirs, Not Latecomers

The cumulative evidence from linguistics, genetics, geography, and history leads to an inescapable conclusion: the Āzrah (Hazara) people are not a "Mongol offshoot" that appeared only after the 13th-century invasions. Rather, they are the direct descendants of the Neolithic Iranian farmers who first settled the central Afghan highlands and who created the Avestan civilization. The Mongol component that is present in their gene pool (haplogroup C3) is a later admixture layer that was added to an already-existing deep substrate. It did not replace that substrate; it blended with it.

The persistent misidentification of the Hazara as "Turco-Mongol" has been driven by three factors:

1. The superficial observation of physical features (epicanthic eye folds, etc.) that were indeed introduced by the Mongol and Turkic migrations. However, these features are the result of later admixture and do not erase the underlying genetic and linguistic continuity.

2. The political narratives of neighboring nationalisms—in Iran, Tajikistan, and among Pashtun nationalists—that have sought to delegitimize the Hazara as "true natives" in order to justify their own territorial and cultural claims. These narratives have repeatedly been shown to be historically and scientifically unsound.

3. The conflation of "Aryan" identity with European racial theories that never applied to the eastern Iranian world. By disentangling the ancient Avestan airya from the modern pseudoscientific "Aryan race," we can see that the Āzrah are the direct bearers of the ancient Aryan identity—not in a racial sense, but in a linguistic, cultural, and territorial sense.

---

7. Conclusion

This paper has presented a multidisciplinary case for the following conclusions:

1. Linguistic: The Hazara self-designation Āzrah is the living continuation of the Avestan term Arəzahī, the eastern clime of light. The Hazaragi dialect preserves over 25–35% of its basic vocabulary directly from Avestan, with no Middle Persian mediation, making it a living fossil of the ancient language.

2. Genetic: The Āzrah carry the highest frequency of the Neolithic Iranian haplogroup J2a1-Page55 (13%) and G2 (8.7%) among all Afghan ethnic groups, demonstrating long-term biological continuity in the central highlands. The later Mongol component (C3) is an admixture layer that does not erase this deeper substrate.

3. Geographical: The Avestan sacred geography—Airyanem Vaejah, Harā Bərəzaitī (Koh-i Baba), and Haētumant (the Helmand River)—is located precisely in the central highlands of Afghanistan, the same region that the Āzrah call home. This identification is supported by the leading Iranologists Michael Witzel (Harvard) and Gherardo Gnoli (Sapienza, Rome).

4. Historical: External sources from the 2nd century CE (Ptolemy) and the 7th century CE (Xuanzang) confirm that this region and its people were identified as "Aryan" long before the Mongol invasions, contradicting the myth that the Hazara appeared only after the 13th century.

5. Conceptual: The ancient Avestan airya (noble, free) is distinct from the 19th-century European pseudoscientific "Aryan race." This study rejects the latter and affirms the former.

On the basis of this evidence, the common historical narrative that reduces the Hazara to a mere Mongol offshoot is scientifically incomplete and incorrect. The Āzrah people are the indigenous heirs of the Avestan civilization. Their name, their language, their genes, and their sacred land all testify to an unbroken chain of

continuity stretching from the first light that rose over Arəzahī to the present day. The history of the Iranian East must be re-evaluated with the Hazarajat as a central, foundational pillar—and the Āzrah as its living voice.

---

8. References

1. Di Cristofaro, J., et al. (2013). Afghan Hindu Kush: Where Eurasian sub-continent gene flows converge. PLOS ONE, 8(10): e76748. DOI: 10.1371/journal.pone.0076748

2. Gnoli, G. (1980). Zoroaster's Time and Homeland. Naples: Istituto universitario orientale.

3. Witzel, M. (2000). The Home of the Aryans. Harvard University. (Published in The Indo-Aryans of Ancient South Asia, ed. G. Erdosy, Berlin: de Gruyter).

4. Xuanzang (tr. Beal, S.). (1884). Buddhist Records of the Western World. London: Trübner.

5. Ptolemy (tr. Stevenson, E. L.). (1932). Geography. New York: New York Public Library.

6. The Avesta: Ābān Yašt, Zamyād Yašt, Mihr Yašt, Vendidad.

7. Bundahišn (tr. Anklesaria, B. T.). (1956). Bombay.

8. Encyclopædia Iranica. "CLIME (kešvar)." "HELMAND RIVER ii. IN ZOROASTRIAN TRADITION."

9. UNESCO (1950, 1951). The Race Question – UNESCO statements.

10. Bartholomae, C. (1904). Altiranisches Wörterbuch. Strassburg: Trübner. (For Avestan etymologies).

11. Morgenstierne, G. (1973). Etymological Vocabulary of Pashto. Oslo. (For comparative Iranian linguistics).

12. Encyclopædia Iranica. "AVESTAN GEOGRAPHY." (Online edition).